

Updated Combined Capability List – December 16, 2011

Note: Reference numbers for the capabilities refer to the original list of 99 from the Charrette and from the virtual sessions, both available at <http://earthcube.ning.com/page/capabilities>

A. Dataset and Workflow Discovery

- A1: Registration of datasets or workflows (with input parameters) into shared collections or global catalogs
- A2: Upload & publish workflows and data sets
- A3: Search across multiple granularity levels & disciplines for workflows and datasets (using semantic metadata and provenance information when appropriate)
- A4: Data subscription services

Summarized from 6, 63, 67, 76, 80, 89, 93, 98, V3, V27, V28

B. Metadata for Workflow and Data Sets

- B1: Automated provenance tracking and tracking of data updates (versioning)
- B2: Commenting, annotating, rating, and categorization of workflows, data sets, or models (both automatic and manual)
- B3: Computational environment provenance stored including code, configurations, input and metadata, with a goal toward reproducibility

Summarized from 11, 14, 21, 22, 27, 43, 47, 68, 69, 70, 73, 74, 82, 86, 88, 95, 97, V1, V6, V9, V26

C. Data Security and Trust

- C1: Provide single sign-on environment for shared collections spanning administrative domains
- C2: Flexible access & sharing controls (licensing) of data, models & workflows
- C3: Issue tracker for problems with data or workflows
- C4: Protect individual property rights
- C5: Data trust: ability for users to track data that needs further explanation or is suspect, fault tolerances, automated tools for validation

Summarized from 16, 18, 57, 64, 75, 76, 84, 92, 96, V7, V20

Closely related to items in F. Data Management within Workflows and M. Policy Enforcement Processes

D. Data Access Services

- D1: Reusable/shared/standard software interfaces for disparate data types
- D2: Brokering interfaces to manage access to data using disparate storage and all protocols
- D3: Deliver data in user-requested format and translation between standards
- D4: Real-time access to data and facilities even in low bandwidth settings
- D5: Networking and linkage of existing data sets
- D6: Data curation: Long term preservation, integrity, authenticity & chain of custody

From capabilities list: 1, 2, 3, 5, 7, 13, 17, 62, 63, 65, 66, 72, V1, V2, V20, V30, V31

Closely related to items in I. Numerical methods and software engineering

E. Workflow Execution Management

- E1: Manage workflow execution in distributed environment – conditional execution, integration with server-side and compute side, interactive
- E2: Combine components from multiple existing workflow systems or models

Summarized from 29, 33, 47, 77, 78, 79, 83, 85, 90, 94, V10, V24, V30

F. Data Management within Workflows

- F1: Caching of intermediate workflow results
- F2: Automatically propagate uncertainty through workflows
- F3: Move workflows to data when more sensible
- F4: Automation of QC/QA procedures where feasible

Summarized from 81, 87, 91, V25

See also C. Data Security and Trust

G. Modeling Standards and Frameworks

- G1: Joint 4D framework for interdisciplinary models, data & information
- G2: Community-based/policed repositories, standards and governance structures for EC compliant tools, and applications that are promulgated
- G3: A means to discover, publish and reuse computational models within the EarthCube framework
- G4: The ability to compose and integrate models or to extend models into scientific workflows

Summarized from capabilities: 25, 26, 28, 34, 35, 37, 42

H. Modeling Capabilities within Cloud, Grid, HPC, and Science Portals

- H1: Web service creation and publishing
- H2: Intelligent data query, retrieval, download and interaction
- H3: Interacting gridding/regridding, visualization and other manipulation and analysis tools

- H4: Real-time simulation capabilities and flexible access to high performance computing

Summarized from capabilities: 30, 31, 36, 39, 40, 41, 45, 46

I. Numerical Methods and Software Engineering

- I1: Experimental facilities and data for model validation
- I2: Automated software building and validation to ensure stable software releases (e.g. the NMI build-test lab)
- I3: Standard APIs and model description standards that enable the creation of better and more reliable tools
- I4: Fault tolerance and reliability built into archival and other systems

Summarized from capabilities: 23, 38, V13, V14, V15, V16

Closely related to items in L. Best Practices and D. Continuity, sustainability, and evolution

J. Tools to Probe, Validate, Verify, and Visualize Data

- J1: 4D integration of high resolution topography scans & geodetic data
- J2: Integration of geologic data in deep time
- J3: Fusion tools that support integration, assimilation & regridding
- J4: Data mining tools and techniques

Summarized from capabilities: 4, 8, 9, 10, 12, 48, 71, 72

K. Broad Participation: Enable Collaboration and Participation from International, Industry, Academic, NGO and other Domain Partners

- K1: Linkage with NEON, LTER, state geological surveys, and other communities
- K2: Low barrier to participation and mechanisms to ensure individual/small voices are heard
- K3: Outreach to encourage new collaborations

Summarized from capabilities: 32, 52, 54, 59, 60, 61, V18, V19, V29, V31

L. Best Practices & Governance Models for the Development of Definitions & Standards

- L1: Community identification & commission of programs of work that are deemed important
- L2: Standards for gridding in models/datasets
- L3: Standards and best practices for formal data publication and citation

Summarized from capabilities: 55, 58, 99, V4, V5, V8

Closely related to items in I. Numerical methods and software engineering

M. Policy Enforcement Processes

- M1: Archival policies for integrity, authenticity, versioning, provenance
- M2: Quality assurance policies

- M3: Role of publishing houses of scientific literature and engagement to drive compliance with community agreed-upon standards
- M4: Role of funding agencies in driving compliance with community agreed-upon standards
- M5: Ensure community consensus
- M6: Consensus driven decision making

Summarized from 5, 15, 19, 20, 22, 43, 50, 53, 56, V22

Closely related to items in C. Data Security and Trust

N. Continuity, Sustainability, & Evolution

- N1: Social networking sites to support knowledge sharing between disparate teams of computer science/geo scientists
- N2: Global catalogs /directories for data, software, models, workflows, etc.
- N3: Linkage with NEON, LTER, state geological surveys, and other communities
- N4: Fault tolerance and reliability built into archival and other systems
- N5: Usage tracking for tools and datasets
- N6: User support services
- N7: Reward systems for all project roles and contributions
- N8: Long-term financial sustainability planning

Summarized from capabilities: 6, 23, 44, 49, 50, 51, 54, V11, V12, V17, V21, V23